

BANDIT MARKET MAKERS

NICOLÁS DELLA PENNA
ANU AND NICTA
ME@NIKETE.COM

MARK D. REID
ANU AND NICTA
MARK.REID@ANU.EDU.AU

ABSTRACT. We propose a flexible framework for profit-seeking market-making, using a sequence of cost-function based automated market-makers with bandit learning algorithms. We do this by considering the magnitude to which a cost-function extends beyond the simplex as a bandit arm, and the minimum-expected profits consistent with a no-arbitrage condition as the rewards. This allows for the creation of market-makers that can adjust bid-asks spreads dynamically, maximising worst-case-expected profits.

1. INTRODUCTION

Motivated by the seminal work of [?], connecting sequential proper-scoring rule based prediction market-makers and no-regret learning with expert advice, we take the first step in examining profit-maximising market-making via a sequence of proper-scoring rules, considered from a no-regret learning perspective. The market-makers traditionally considered in these prediction market context have instantaneous asset prices that sum to one; that is to say, the prices lie on the simplex. This allows for the prices to be interpreted as probabilities, and when combined with path-independence, implies that the market-maker cannot be profitable in expectation [?]. We relax both. We consider a notion of regret relative to a market-maker that at any point in time presents a proper-scoring rule within a certain class, and uses a no-arbitrage condition to price this change in a position that corresponds to that parametrisation of the proper-scoring rule. We show that under this setting the problem can be reduced to a suitably constructed continuous-armed bandit problem, using standard algorithms and analysis. In particular, a no-arbitrage assumption on the prices at the end of every time-period implies that the probability of an even-occurring is bounded by the bid-ask. This is a natural and common assumption in finance, which to the best of our knowledge has not previously been explored in the sequential proper-scoring rule setting.

In our reduction to the market setting, each action in the bandit setting corresponds to a parametrisation of a cost-function. The rewards on the bandit setting correspond to minimal-expected profits for the trades incurred in that period that are consistent with the final set of prices in that period. The bandit algorithm provides a way to dynamically adjust the sum of prices of assets so as to asymptotically extract maximal minimum-expected profits for the class of cost-function under consideration.

Given our use of the underlying no-regret bandit algorithms, we are able to obtain regret bounds on the sum of the minimal per-period-expected profits of our market-maker. These regret bounds are, however, limited in two ways. First, they do not refer to the actual profits of the market-maker, but rather to the sum of the minimal-expected profits at each time-period that is consistent with market prices. Second, the comparison class is not any market-maker, but rather market-makers that use a specific family of cost-functions. These two limitations appear to be more general than our particular reduction, and cannot be avoided without stronger assumptions. In particular, our analysis assumes nothing about the distribution of traders' beliefs and the demands they generate beyond a very weak and standard smoothness restrictions. Stronger assumptions are needed to estimate down-expected profits at each time-period more precisely; for example, the way [?] leverages an assumption of symmetry around the truth of traders' beliefs). Secondly, the use of a sequence of sequentially-shared proper-scoring rules allows us to provide worst-case boundaries on how much the market-maker can lose that are model free (to the extent that we are aware, this dual guarantee both in terms of profits relative to the best market-maker in the class with a model assumption, and also of a fixed worst-case loss even when the model assumptions do not hold); this too is a novel approach (for example in both [?] the worst-case loss if assumptions are wrong are unbounded).

Our paper is structured as follows: In §2 we survey related work and introduce some motivating examples in §3, before introducing our bandit market-making framework in §4. Our main result—a regret bound on worst-case maximum-expected profits for our market-maker—is given in §5, before we close with a discussion of possible future work in §6.

2. RELATED LITERATURE

Two streams of literature exist on market-making. One is focused on eliciting information (prediction markets), and the other is focused on making a profit by providing liquidity to traders, motivated to trade for exogenous reasons such as hedging. To our knowledge, these have largely remained separate.

The market-maker as a way to elicit information from traders has been studied in an extensive literature on cost-function based automated market-makers motivated by prediction markets [? ? ?]. By equating outcomes in

the market setting with experts in the learning setting, and the trades made in the market with the losses observed by the learning algorithm Chen et al [?] establish the striking mathematical equivalence between cost-function based prediction markets and regularised follow the leader online learning. The learning that takes place in these cases can be seen to be taking place over the probabilities of the events. In these cases, the market-maker can be seen as subsidising trade in securities contingent on an outcome onto which traders might not have exogenous reasons to hedge, as a way of paying to extract their beliefs about the likelihood of the events.

Cost-function based market-makers for prediction markets are based on sequentially shared proper-scoring rules. These are myopically incentive compatible: that is, if traders do not consider the effects of their trades on other players beliefs or on the market-makers future actions then proper-scoring rules incentivise players to reveal their true beliefs. If traders can interact multiple times with the market-maker and act strategically, proper scoring rules are not enough to incentivise traders reveal their true beliefs [?].

likewise, the market-maker as an agent motivated to make profits by offering liquidity to traders and attempting to maximise profit has been the subject of an extensive literature [? ? ?]. Learning can also be considered to be taking place in this case, but the learning is over which bid-ask schedules will balance the supply and demand and extract the largest profits. Cost-function based automated market-maker with prices that sum to greater than one have been studied in [? ? ?] which implies these market-makers that can turn a profit. These market-makers do not, however, optimise the amount of profits they extract from traders.

Using a bandit framework to attack a pricing problem goes back in the economics literature to [?] who considers the case with discounting. Special cases of the problem have more recently been analysed using modern methods in [?]. A specific version of online posted-price auctions has been analysed using adversarial bandits in [?]. An aspect that distinguishes the for-profit market-maker's problem from that of a seller or auctioneer is that the market-maker does not know its cost for the assets sold, but rather must estimate it based on the prices at which the market clears.

3. MOTIVATING EXAMPLE

To clarify the role of the sum of prices (a.k.a. the *overround* in the framework denoted by a) we consider the following toyexample. The scenario can be represented by a table containing normalised-expected rewards per-round to the market-maker. Rows represent the distribution of trader demand, and columns the market-maker's choices. The distribution of trader demand is unknown to the market-maker at the start of the game and may vary over time.

	low a	high a
Low variance	0.45	0
High variance	0.45	1

We consider a case where the true probabilities of the outcome are known to the market-maker, and the traders are split between two possible underlying distributions of beliefs: one that** is tightly concentrated around the truth (low variance), and one that is spread out (high variance). This can be considered as a situation where there is more disagreement between traders (and the market-maker can potentially extract high profits) and one where there is less disagreement (and thus, less potential profits for the market-maker overall).

A high value of a means the market-maker makes a smaller number of trades, but each is at a higher margin. In this case, it is advantageous when beliefs have higher variance, but the low variance situation can lead to no trades occurring, since the entire distribution of beliefs can fall between the price of the asset and the price of its complement (i.e. the bid-ask spread). In this low variance situation we say the traders cannot *profitably express* their beliefs and formalise this notion in Section 4.

If the Exp3 bandit algorithm [?] was used to make choices using our framework in this example, the market-maker would converge (with high probability) to playing whichever choice has the highest-average payoff. Now suppose all high-variance in belief states were to occur in the initial periods, and all low-variance in beliefs states in the later periods. This would result in the market-maker choosing the high a at the beginning and then switching to the low a at the end. In Section 5, we use a regret-bound result for Exp3 which provides convergence guarantees against classes of competing strategies (in this case, sequences of choices of a) that can make a fixed number of switches.

4. FRAMEWORK

We now give a high level overview of our bandit market-making framework. Details and assumptions about the various components are given in the subsections below.

We consider a setting with n mutually exclusive and collectively exhaustive outcomes, and T time-periods of trading. At each time-period $t = 1, \dots, T$ the market has an obligation vector q_t and a bandit algorithm selects an “arm” in the form of a cost-function C from a predefined set \mathcal{C} which defines the market-maker for round t . Traders interact with the market-maker by buying portfolios of contracts at prices set by the market-maker. The aggregate purchases of the sequence of traders that arrives at period t shifts the obligation vector q from its previous state q_{t-1} to q_t , the quantity vector at the end of the period.

It is important to note that our worst-case loss is bounded by the worst-case loss of the cost-function that provides the most liquidity in the set under consideration.

4.1. Cost-functions. The behaviour of a market-maker is defined through its *cost-function* $C : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$. This function assigns a monetary value $C(q)$ to each *market position* described by a vector $q \in Q$ where Q is some bounded subset of \mathbb{R}_+^n . Each component q_i is the total size of the obligation vector in case event i occurs. If the market is in position q and a trader wants to buy a portfolio of r shares, the price the trader must pay is $C(q+r) - C(q)$. This means the *instantaneous price* per-share for each security i is $\frac{\partial}{\partial q_i} C(q)$ and can be summarised by the vector $\pi(q) := \left(\frac{\partial C(q)}{\partial q_1}, \dots, \frac{\partial C(q)}{\partial q_n} \right)$; that is, the gradient $\nabla C(q)$. We will sometimes denote prices π without the argument when the obligation vector is clear from context.

We put some natural restrictions on the cost-function, similar to those used in [? ?]:

- **Convexity:** $C(\lambda q + (1-\lambda)q') \leq \lambda C(q) + (1-\lambda)C(q')$ for all $q, q' \in Q$ and all $\lambda \in [0, 1]$.
- **Monotonicity:** $C(q) \geq C(q')$ for all $q, q' \in Q$ such that $q \geq q'$ (that is $q_i \geq q'_i$ for $i = 1 \dots n$).
- **Bounded Loss:** $\sup_{q \in Q} \max_i q_i - C(q) < \infty$.

In addition we require that the cost-function always offer prices that are *potentially profitable for the market-maker*:

- **Potentially Profitable:** $\sum_{i=1}^n \pi_i(q) \geq 1$ where for all $q \in Q$, and $\sum_{i=1}^n \pi_i(q) > 1$ for some q .

The potentially profitable condition requires that prices sum to greater than one for for some q and never to less than one. That is, for some state of the quantity vector, the cost-function should offer a set of prices that if a uniform vector of assets is purchased by traders it will make a profit with certainty. When prices sum to less than one, it opens the market to risk-free arbitrage opportunities on the part of traders. Here we follow [?], and consider complete-market-makers whose prices sum to greater than one, and only allow trades to purchase assets so as to avoid risk-free arbitrage on the part of traders. We will focus on a complete-market setting, so that this prohibition on asset sales is without loss of generality: if traders wish to take positions against a given outcome, they can still do so by buying the basket of assets that make up its complement.

An immediate consequence of a market-maker using a potentially profitable cost function C is that there are beliefs (subjective probabilities over the outcomes) for traders which are not *profitably expressible*. That is, if a trader believes the outcome probabilities are $p \in \Delta^n$ and the market's obligation vector is q , there is no portfolio $r \in \mathbb{R}_+^n$ that can be bought so that $C(q+r) - C(q) \leq \sum_i p_i r_i$. We use Δ_π^n to denote the subset of *compatible*

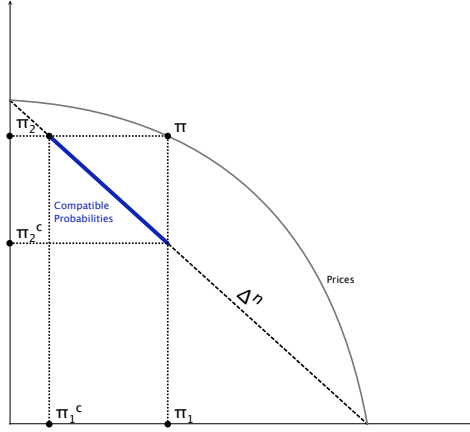


FIGURE 1. Compatible probabilities on the simplex Δ^n for the prices π .

probabilities for price π . Letting $\pi_i^c = 1 - \sum_{j \neq i} \pi_j$ we define

$$\Delta_\pi^n := \{p \in \Delta^n : p_i \in [\pi_i^c, \pi_i] \text{ for all } i = 1, \dots, n\}$$

Given some vector q , there exists a vector of beliefs over outcomes p , such that a trader who believes that the true probability distribution is p cannot engage in any trades and expect to profit. In particular, note that by construction since $\sum_i \pi_i > 1$ there must exist a vector p , such that $1 - \sum_{j \neq i} \pi_j < p_i < \pi_i$ for all i . Intuitively, this region of non-profitable trading means there is a “bid-ask spread” on the prices posted by the market.

Finally, for convenience, we also assume that prices derived from cost-functions satisfy $\pi(q) \in [0, 1]^n$ for all $q \in Q$. Note that since prices are just the gradient of the cost-function, we are implicitly assuming that all cost-functions are Lipschitz with constant $L \leq \sqrt{n}$.

Families of potentially profitable cost-functions can easily be constructed from a given cost-function C with fair prices, by charging a fixed multiple $a > 1$ of the prices for C . Other transformations that preserve properness can be used, such as adding per-share charges, and combinations of the two. We term these *overround cost-functions*. For example, if $C(q) = b \log \sum_i \exp(q_i/b)$ is the well-known logarithmic market scoring rule (LMSR) [?] with fixed parameter b , we can construct the single parameter family $\mathcal{C} = \{aC : a \in [1, \infty)\}$. In this case, the prices charged by some $C = aC \in \mathcal{C}$ at market position q are simply

$$\pi_i(q) = a \frac{\exp(q_i/b)}{\sum_i \exp(q_i/b)}.$$

We will refer to this family of potentially profitable cost-functions built upon the LMSR as the *overround LMSR* cost-functions.

When treating a set of cost-functions as arms in a bandit game in Section 5, we require a metric measure of “closeness” between cost functions. The most convenient measure for our purposes is that which takes the maximal difference in costs assigned to any market position by two given cost-functions. Formally, we define the metric

$$d_\infty(C, C') := \sup_{q \in Q} |C(q) - C'(q)|$$

and consider two cost-functions to be close if $d_\infty(C, C')$ is small. It is easy to establish that d_∞ is indeed a (pseudo-)metric: $d_\infty(C, C') = d_\infty(C', C) \geq 0$ with $d_\infty(C, C) = 0$ and the triangle inequality carries over from the triangle inequality of $|x - x'|$ on the reals.

4.2. Traders. At each time-period t , a finite number of traders have an opportunity to trade with the market-maker in sequence. Traders’ demands for the assets at given prices are drawn from a distribution D_t . We follow the prediction markets literature and consider how traders interact with the market-maker in a myopic manner, which is sufficient to avoid question-of-incentive compatibility, both on the cost-function and the bandit algorithm used.

For any choice of C and q , a belief distribution D induces a random variable $\bar{q} \in Q$ of final market state, after trading with a finite number of traders with demands drawn using D .

Since the reward for a cost-function depends on the random state vector at the end of the trading round, and this in turn is dependent on the choice of cost function made by the market at the beginning of the round, we need some assumption as to how much a small change in the latter can affect the former. We define $\bar{q}(C)$ to be the random state vector after traders drawn from D interact with the market-posting prices according to C . Each trader has beliefs, utility, and a bounded budget, and trade myopically to maximise their expected utility. Given a metric d defining the “closeness” of any two cost functions, we assume that D and these properties of the traders are such that

$$\bar{q}(C) - \bar{q}(C') \leq K d(C, C')$$

for some fixed constant K .

This can be interpreted as giving the same starting quantity vector and two cost-functions that are close, the quantity vectors after traders interact with those two cost-functions will also be close. In other words, faced with almost the same incentives, the same traders will respond in almost the same way.

When the above property holds, we say that D is a *d-regular* belief distribution.

4.3. Rewards. A natural measure for the rewards of the bandit for period t to be the expected profits/losses incurred by the market-maker during

the period. The existence of a range of beliefs that are not profitably expressible implies that there is no unique way to transform the price vector into a probability vector; thus we consider the expectation with regards to the worst-case compatible probability distribution over outcomes (for the market-maker) consistent with the prices of the quantity vector at the end of the period.

Thus at each time-period our bandit's unnormalised rewards are given by:

$$(1) \quad r_q(C) = \overbrace{C(\bar{q}) - C(q)}^{\text{Market Earnings}} - \overbrace{\max_{p \in \Delta_\pi^n} \sum_{i=1}^n p_i(\bar{q}_i - q_i)}^{\text{worst-expected payout}}$$

For any choice of C we see that for all $q \in Q$ we have $r_q(C) \leq C(\bar{q}) \leq \sup_{q' \in Q} C(q')$ since both $C(q)$ and $\sum_i p_i(\bar{q}_i - q_i)$ are non-negative. The reason the latter is non-negative is that, since traders can only buy contracts, $\bar{q}_i \geq q_i$ for all i . The reward function is also bounded below for any choice of C by $-\sup_{i, q' \in Q} q'_i$ since $C(\bar{q}) - C(q) \geq 0$ by the monotonicity assumption and since $p_i(\bar{q}_i - q_i) \leq \bar{q}_i \leq \sup_{q' \in Q} q'_i$. This lower bound is finite, since we assume Q is bounded. These bounds are used in Section 5 below to normalise the rewards to $[0, 1]$.

Note that our choice of reward function makes us particularly sensitive to the choice of time-period. To illustrate this, consider a market on the outcome of a (possibly biased) coin toss. If on every odd time-period all traders we interact with hold the belief that heads will occur with certainty, and in even time-periods that tails will occur with certainty, all asset purchases in each time-period would be a single asset, and our worst-case-expected loss would be that the event of the asset that was bought will occur with high probability. Thus, while it would be extremely profitable to be a market-maker in this situation that provided liquidity across time-periods, our framework would lead to very low amounts of liquidity if possible, or sufficiently high prices ($a > 2$) to effectively walk out of the market. By making the time-periods twice as long, and thus in each period interacting with a balanced mix of traders on each side (the half from each of the previous even and odd periods), the bandit market-maker would however obtain the optimal profits in its class. This highlights the importance of choosing the length of time-periods to be long enough that the market-clearing price can be reached within them. There is of course a tension here, as the longer the individual time-periods of the bandit algorithms, the fewer the opportunities to exploit the information learned that have a for a fixed amount of potential trades.

4.4. Bandit Algorithms. Due to the fact that the rewards for period t depend on the quantity vector that resulted from the previous periods trading q_{t-1} , it is not possible for these to be considered as having been set by

an oblivious adversary before the market starts. This restricts our choice of bandit algorithms if we wish to have a regret analysis that can cope with adaptive adversaries.

To set the overround and the liquidity simultaneously, we face a bandit problem with two continuous arms. To the best of our knowledge, there are no bandit algorithms for the multidimensional action space for which regret bounds have been obtained under adaptive adversaries. We can, however, follow the prediction markets literature, pick our liquidity *b a priori*, and then use the continuous-arm algorithms of Kleinberg [?] to obtain a worst-case guarantee.

An alternative which does not achieve vanishing regret asymptotically, but rather one that approaches a constant fraction of the optimal profits, is to use a discretisation over the space of arms. How large a fraction is achieved will then depend on the smoothness of the profits with respect to the parameters and the fineness of the discretisation grid (i.e. how much higher are the profits under the true optimal parametrisation than those of the best parametrisation that is part of our discretised set). However, in this parameter-discretised setting, there are extremely attractive bandit algorithms. We can use the classic Exp3 [?], which imposes no stochastic assumption; that is, it allows for an arbitrary sequence of $D_t \neq D_{t-1}$ in the market and obtains a $O(T^{3/4})$ regret bound.

It is important to note once again that we are assuming traders being myopic. If they are not, then the bandit algorithms are not generally truthful [?].

5. REGRET BOUNDS

The aim of the bandit market-maker is to minimise its regret—i.e., the difference between its profits and the profits of the best single choice of cost-function in hindsight.

5.1. Finite Sets of Cost-functions. We must use bandit algorithms for adaptive adversaries, since after each round of trading the rewards shift due to a change in q . To use the Exp3 results, all we need to do is select some number of arms K , then apply the bounds from Theorem 8.1 and corollary 8.2 in [?], which is stated in terms of the *hardness* $H(s)$ of a strategy s (number of action changes over T rounds). Here the regret bound is $H(s)\sqrt{KT \ln(KT)} + 2e\sqrt{\frac{KT}{\ln(KT)}}$.

5.2. Continuous Cost Spaces. To avoid the artificiality of choosing K arms, we would like to make use of bandit algorithms for continuous spaces (e.g., [?]). Without the reward function having some structure over the set of possible arm choices, no regret bound is possible.

Theorem 1.4 in [?] gives a continuous arm bandit result for adaptive adversaries for a smooth class reward functions on a bounded subset S of \mathbb{R} . Before stating his result, we first define the class of reward functions under

consideration. A function $f : S \rightarrow \mathbb{R}$ is said to be *uniformly locally Lipschitz* with constant $L > 0$, exponent $\alpha \in (0, 1]$, and restriction $\delta > 0$ whenever f satisfies, for all $u, u' \in S$ with $|u - u'| \leq \delta$, $|f(u) - f(u')| \leq L|u - u'|^\alpha$. Following Kleinberg, we denote this class of functions $ulL(\alpha, L, \delta)$.

Theorem 5.1 ([?]). *Let S be a bounded subset of \mathbb{R} and $\Gamma \subset ulL(\alpha, L, \delta)$ be a set of uniformly locally Lipschitz reward functions $r : S \rightarrow [0, 1]$. Then there exists an algorithm CAB that achieves regret $O\left(T^{\frac{\alpha+1}{2\alpha+1}} \log^{\frac{\alpha}{2\alpha+1}} T\right)$ against adaptive adversaries.*

Since S must be a subset of \mathbb{R} , we cannot directly use cost-functions from an arbitrary class \mathcal{C} as bandit arms in the above result. Instead, we focus on establishing a regret bound for the single parameter family $\mathcal{C}_{\text{over}}$ of overround cost-functions defined in Section 4.1. A choice of arm $C \in \mathcal{C}_{\text{over}}$ is now equivalent to a choice of parameter $a \geq 1$. To meet the boundedness condition on S , we further restrict the overround class to cost-functions with $a < A$ for some fixed choice of upper bound A . We note that since prices are assumed to be in $[0, 1]^n$ and prices for aC are a scaling by a of prices in the simplex, we see that for valid cost-functions $A \leq \sqrt{n}$.

With the above assumptions in place, the following result states that it is possible for a bandit market-maker to achieve vanishing-regret relative to a single choice of overround cost-function, given the previously discussed regularity assumptions about traders' demands. This means, despite not knowing the distribution of traders' demands, it is possible for the bandit market-maker to asymptotically extract a comparable amount of expected profit, as though the best cost-function in the class for traders with those demands was known in advance.

Theorem 5.2. *Let $A > 1$ be fixed and let $\mathcal{C}_{\text{over}} = \{aC : a \in [1, A]\}$ be the family of overround cost-functions with metric $d_\infty(C, C') := \sup_{q \in Q} |C(q) - C'(q)|$. Then there exists a bandit market-maker (using the CAB algorithm) that achieves regret $O(T^{2/3} \log^{1/3} T)$ against traders with beliefs drawn according to d_∞ -regular distribution D_t with a common regularity constant K .*

The proof of this theorem hinges on showing that the reward functions r_q can be transformed into rewards for a bandit game that is $[0, 1]$ -valued and uniformly locally Lipschitz with $\alpha = 1$. Once this is done, the algorithm CAB can be used to choose arms $a \in [1, A]$ corresponding to cost-functions $C = aC \in \mathcal{C}_{\text{over}}$. The regret bound for the game over $[1, A]$ then carries over to the game on $\mathcal{C}_{\text{over}}$.

Proof. Fix $q \in Q$ and define the function $R_q(a) := \frac{1}{AM}(r_q(aC) - m)$ where r_q is defined in equation (1), $M = \sup_{q' \in Q} |C(q')|$, and $m = -\sup_{i, q' \in Q} q'_i$. This function $R_q(a)$ is bounded with $[0, 1]$ since $r_q(C) \leq m$ independent of C , as per the discussion in Section 4.3, and because $r_q(aC) \leq ar_q(C)$. Thus, $\frac{r_q(aC) - m}{AM} \leq \frac{a}{A} \frac{r_q(C)}{M} \leq 1$.

Now let $C, C' \in \mathcal{C}_{\text{over}}$ so that $C = aC$ and $C' = a'C$ with $|a - a'| < \delta$. Let $\pi = \nabla C(q)$ and $\pi' = \nabla C'(q)$ be the prices for C and C' at q . Let \bar{q} and \bar{q}' denote the final market position after the traders drawn from D interact with markets starting at q with cost-functions C and C' , respectively.

Note that the definition of the overround cost-function means that

$$(2) \quad d_{\infty}(C, C') = \sup_q |C(q) - C'(q)| = M |a - a'|$$

where $M = \sup_q |C(q)|$. Now observe that

$$\begin{aligned} |r_q(C) - r_q(C')| &= \left| C(\bar{q}) - C(q) - \max_{p \in \Delta_{\pi}^n} \sum_i p_i(\bar{q}_i - q_i) - C'(\bar{q}') + C'(q) + \max_{p' \in \Delta_{\pi'}^n} \sum_i p'_i(\bar{q}'_i - q_i) \right| \\ &\leq \underbrace{|C(\bar{q}) - C'(\bar{q}')|}_{T_1} + \underbrace{|C(q) - C'(q)|}_{T_2} + \underbrace{\left| \max_p \sum_i p_i(\bar{q} - q) - \max_{p'} \sum_i p'_i(\bar{q}' - q) \right|}_{T_3}. \end{aligned}$$

The term T_2 is clearly bounded by $d_{\infty}(C, C')$ and thus $T_2 \leq M |a - a'|$ by (2). The term T_1 can be bounded by noting that

$$\begin{aligned} T_1 &= |C(\bar{q}) - C(\bar{q}') + C(\bar{q}') - C'(\bar{q}')| \\ &\leq |C(\bar{q}) - C(\bar{q}')| + |C(\bar{q}') - C'(\bar{q}')| \\ &\leq L \|\bar{q} - \bar{q}'\| + M |a - a'| \end{aligned}$$

since C has Lipschitz constant $L \leq \sqrt{n}$, as discussed in Section 4.1.

Note that T_3 is the difference between two optimisations: one over $p \in \Delta_{\pi}^n$, and the other over $p' \in \Delta_{\pi'}^n$. Observe that because C and C' are overround cost-functions, their prices at q are $\pi = a\hat{\pi}$ and $\pi' = a'\hat{\pi}$ where $\hat{\pi} \in \Delta^n$ is the prices given by C . Thus, π and π' lie on the same ray through Δ^n , and therefore either $\Delta_{\pi}^n \subset \Delta_{\pi'}^n$, or *vice versa*, depending on whether $a < a'$ or not. Letting p^* be the maximising point within the smaller region gives $T_3 \leq |\sum_i p_i^*(\bar{q}_i - q_i) - \sum_i p_i^*(\bar{q}'_i - q_i)| = |\sum_i p_i^*(\bar{q}_i - \bar{q}'_i)|$, and so $T_3 \leq \|p^*\| \|\bar{q} - \bar{q}'\| \leq \|\bar{q} - \bar{q}'\|$ by the Cauchy-Schwarz inequality, and the fact that $\|p^*\| \leq 1$ as it is on the simplex.

Thus, we have that

$$|r_q(C) - r_q(C')| \leq (L + 1) \|\bar{q} - \bar{q}'\| + 2M |a - a'|$$

but because, by assumption, D is d_{∞} -regular $\|\bar{q} - \bar{q}'\| \leq K d_{\infty}(C, C') = K M |a - a'|$ and so $|r_q(C) - r_q(C')| \leq ((L + 1)K + 2)M |a - a'|$ and thus R_q is uniformly locally Lipschitz with constant $\frac{(L+1)K+2}{A}$ and $\alpha = 1$, as required. \square

We briefly note the Lipschitz constant for R_q is $O(\sqrt{n}K)$, since $L \leq \sqrt{n}$ and $A \geq 1$. As the Lipschitz constant for the reward appears multiplicative in Kleinberg's regret bound, both the number of outcomes n and the degree

of regularity K have an adverse impact on the convergence of the bound. This aligns with intuition, since K measures the volatility of final market states as a function of C .

6. DISCUSSION AND CONCLUSIONS

Given our choice of reward function to obtain a regret guarantee, an adaptive adversary is essential, since the reward function necessarily varies after each round of trading due to a change in q . Some assumptions about the influence of the demand distribution D on the final q as the cost-function varies seems necessary too. Wildly varying responses to small changes in cost-functions would make any learning impossible.

Several avenues for future research, extending the results that can be obtained in the bandit market-making framework, are of substantial interest. We have only established a regret bound for overround cost-functions with respect to the sub-norm metric, leaving much room for extensions. For example, it should be possible to exploit the convexity, etc., of cost-functions to have the result hold under less extreme norms (e.g., L^1), and for different 1-parameter families, such as p -ball cost-functions [?].